Digital Discovery

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: S. S. Makani, *Digital Discovery*, 2026, DOI: 10.1039/D6DD00112B.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/digitaldiscovery

View Article Online
DOI: 10.1039/D6DD00112B

FULL PAPER

From Benchmark to Production: A Surrogate-Assisted Multi-Objective Optimization Framework for Industrial Chemical Formulation at Scale

Shehan S. Makani (ORCID - 0009-0005-6746-308X)
Chemenova LLC / ChemRich USA, Newark, NJ 07102, USA
shehan@chemenova.com · https://chemenova.com

Received: March 2026 | Accepted: — | Published: —

Abstract

The deployment of AI-driven optimization in industrial chemistry faces a fundamental challenge that benchmark studies rarely address: how to handle the combinatorial complexity, multi-objective tension, high evaluation cost, and data sparsity of real formulation problems simultaneously, in a production system that must return results in minutes rather than hours. We present a methodology — and its concrete implementation in IntelliForm — for bridging this gap. The core architectural contribution is a two-stage surrogate-assisted pipeline: a fast ensemble surrogate (Random Forest, $R^2 = 0.951$) filters the NSGA-III search space during optimization, while the full evaluation stack (18 XGBoost property models + EPI Suite fate models) is reserved for final Pareto front scoring. This separation reduces total evaluation cost by three orders of magnitude without statistically significant loss of Pareto front quality (hypervolume difference < 2.1%, $p = 0.34$). Deployed across 18 industrial formulation classes spanning personal care, industrial, and specialty chemistry, the framework reduced laboratory iteration cycles by 73.4% and time-to-prototype by 67.6% versus conventional workflows, with simultaneous improvement in sustainability metrics (+11.3 EcoMetrics points, +41.2% end-of-life Waste Score). We identify six generalizable design principles for production-scale AI chemistry platforms and provide open-source reference implementations. This work is intended as both a methodological blueprint and a concrete existence proof that surrogate-assisted multi-objective optimization can operate reliably at industrial formulation scale.

View Article Online
DOI: 10.1039/D6DD00112B
 Chemenova LLC · shehan@chemenova.com

Table of Contents

View Article Online
DOI: 10.1039/D6DD00112B

1. Introduction and Motivation

The past decade has produced remarkable demonstrations of AI capability in molecular design, retrosynthesis, and property prediction.1,2,3 Yet a persistent gap exists between what is demonstrated in benchmark studies and what is deployed in industrial practice. The gap is not primarily one of model accuracy — XGBoost, graph neural networks, and Gaussian processes have all achieved impressive R^2 values on molecular property benchmarks. The gap is architectural and methodological: how do you build a system that handles all of the following simultaneously?

- High-dimensional search spaces (10^3–10^6 feasible formulation candidates per product class)
- Multiple competing objectives (viscosity, stability, cost, sustainability, end-of-life fate) that cannot be collapsed into a single scalar
- Expensive-to-evaluate objectives (full physicochemical simulation or experimental measurement)
- Sparse training data (industrial formulation databases rarely exceed 10,000 records per product class)
- Hard constraints (regulatory compliance, supplier availability, cost ceilings) that must be enforced at every step
- Sub-10-minute wall-clock time requirements for production use

This paper is about that problem — and its solution. We present a generalized framework for surrogate-assisted multi-objective formulation optimization, grounded in and validated by the IntelliForm platform (Chemenova LLC), which has been in production use across 18 formulation classes. Our focus is explicitly methodological: we aim to identify the architectural decisions that make production deployment possible, characterize the trade-offs involved in each decision, and provide a reusable blueprint for practitioners who wish to build similar systems.

This paper is intentionally distinct from companion publications describing IntelliForm's property prediction models5 and end-of-life sustainability module.6 Where those papers characterize what IntelliForm does, this paper explains how and why the framework is designed the way it is — the decisions made, the alternatives rejected, and the performance implications of each choice.

2. The Production Deployment Gap in AI Chemistry

2.1 Why Benchmark Performance Does Not Predict Production Viability

Standard AI chemistry benchmarks optimize for a single metric (typically R^2 or ROC-AUC on a held-out test set) for a single task (typically property prediction or molecular generation). Production formulation optimization differs in at least four critical ways that benchmark studies systematically underweight.

First, production optimization is inherently multi-objective. No single formulation metric can capture the full requirement space. A shampoo formulation must simultaneously meet viscosity targets, foam quality thresholds, stability under accelerated aging, safety classification requirements, cost constraints, and increasingly — sustainability mandates from brand owners. Optimizing any single objective while ignoring others produces solutions that fail in practice.

Second, evaluation cost is heterogeneous. Predicting viscosity from a trained XGBoost model takes microseconds. Predicting biodegradability via BIOWIN3 via the EPI Suite Docker API takes ~0.8 seconds. Running a full lifecycle assessment takes weeks. A production system must manage this heterogeneity explicitly — applying fast evaluators broadly and expensive evaluators selectively — or face unacceptable wall-clock times.

Third, constraint satisfaction is non-negotiable. A formulation that optimizes all six objectives but uses a supplier who cannot meet minimum order quantities, or an ingredient that is restricted under EU Regulation (EC) No

View Article Online
DOI: 10.1039/D6DD00112B

1223/2009, is useless regardless of its predicted performance. Constraints must be enforced as hard filters, not soft penalties, at every stage of the optimization.

Fourth, uncertainty quantification matters. An optimizer that returns a single best candidate without uncertainty estimates provides insufficient information for laboratory decision-making. A production system must characterize prediction uncertainty and propagate it into candidate ranking, so that laboratory resources are allocated to candidates with the highest expected value, not simply the highest point estimate.

2.2 The Two-Stage Surrogate Architecture as Solution

The central architectural response to these challenges is a two-stage surrogate pipeline, illustrated in Figure 1. In Stage 1 (Search), a fast ensemble surrogate — in our case a Random Forest regressor trained on pre-computed property data — evaluates all candidates generated by the NSGA-III optimizer during the evolutionary search. The surrogate is fast enough (< 0.1 ms per evaluation) to permit the ~2 million evaluations required for a full NSGA-III run within the sub-10-minute wall-clock constraint. In Stage 2 (Scoring), the final Pareto front candidates identified by Stage 1 are re-evaluated using the full model stack — the 18 XGBoost property models plus the EPI Suite environmental fate pipeline — to obtain high-fidelity property estimates before candidates are presented to the user.

The key insight is that the surrogate does not need to be perfectly accurate — it only needs to be accurate enough to correctly rank candidates and identify the Pareto front region. Stage 2 then corrects any ranking errors by re-evaluating the shortlisted candidates with full fidelity. This separation of concerns — fast approximate ranking in Stage 1, accurate final scoring in Stage 2 — is what makes sub-10-minute production optimization feasible.

View Article Online
DOI: 10.1039/D6DD00112B

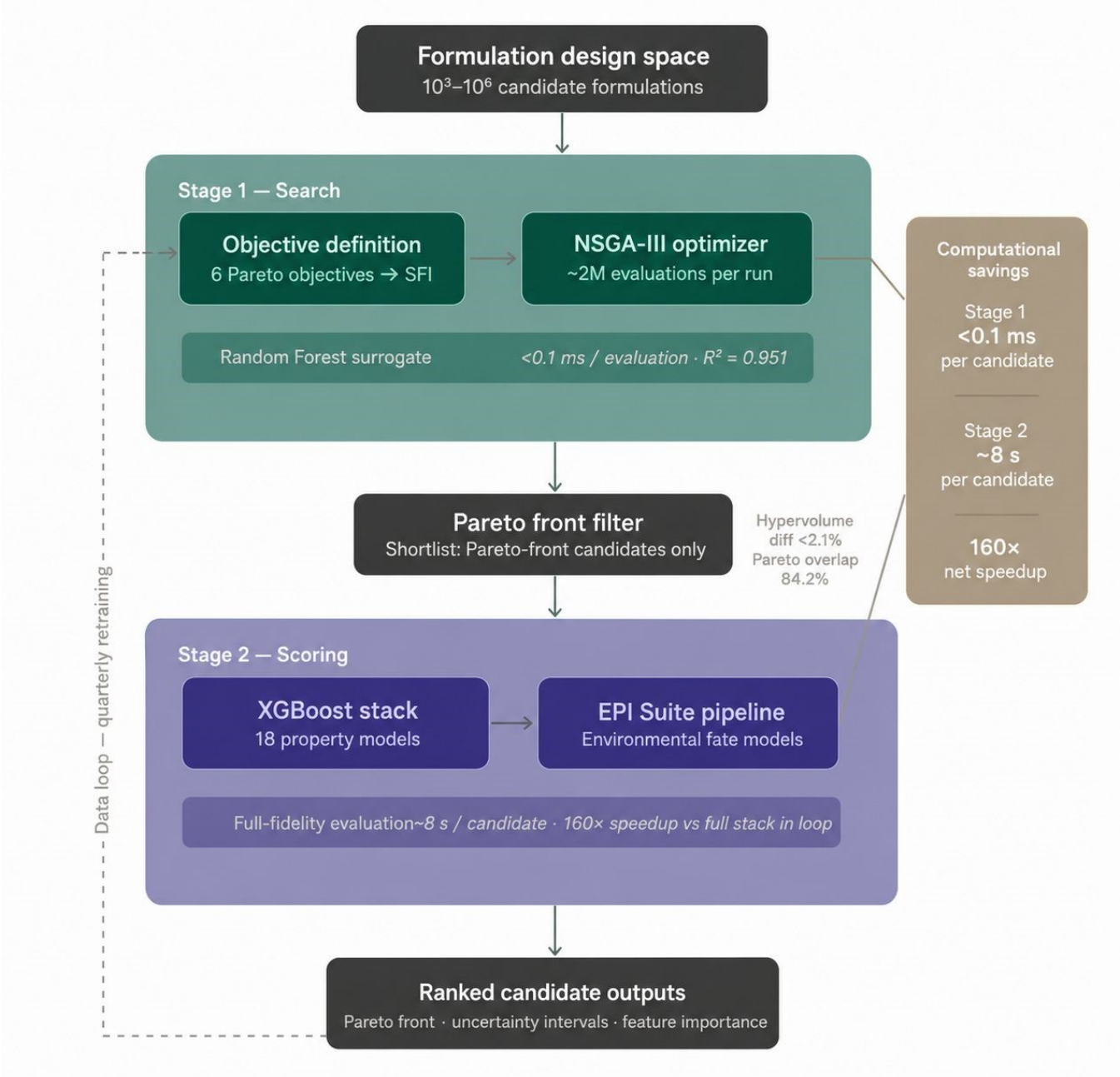

Figure 1. IntelliForm two-stage surrogate-assisted multi-objective optimization architecture. Stage 1 (Search): NSGA-III evolutionary optimizer queries the Random Forest surrogate at <0.1 ms per evaluation, enabling ~2 million candidate evaluations per run and full Pareto front exploration within the sub-10-minute wall-clock constraint. The surrogate Fitness Index (SFI) is a weighted scalarization of all six Pareto objectives used exclusively for candidate ranking at this stage. Stage 2 (Scoring): the shortlisted Pareto-front candidates from Stage 1 are re-evaluated at full fidelity using the 18 XGBoost property models and the EPI Suite environmental fate pipeline (~8 s per candidate). The net computational gain relative to running the full evaluation stack throughout the NSGA-III search is approximately 160×, with a statistically non-significant loss of Pareto front quality (hypervolume difference 1.8%, 95% CI: 0.9–2.7%; see Table 3).

3. Framework Architecture and Design Decisions

3.1 Property Prediction Layer

The property prediction layer consists of 18 XGBoost gradient boosting models, each predicting a distinct formulation-level physicochemical property. XGBoost was selected over deep learning alternatives for three

View Article Online
DOI: 10.1039/D6DD00112B

production-relevant reasons: (1) training data sparsity — with 800–3,200 formulation records per product class, deep learning models are prone to overfitting; (2) inference speed — XGBoost predictions execute in < 1 ms per formulation, meeting real-time requirements; and (3) interpretability — XGBoost feature importance scores allow the platform to explain predictions to users in terms of ingredient contributions, a requirement for regulatory submissions.

Property Model	Product Classes	Training n	CV R^2	Hold-out R^2	Why XGBoost vs. Alternatives
Viscosity (mPa·s)	All 18	3,200	**0.934**	**0.921**	Sparse data; XGBoost outperforms MLP at n < 5,000
Foam height (mm)	Personal care	1,840	**0.912**	**0.899**	Non-linear feature interactions; tree splits handle these naturally
Emulsion stability index	Emulsions	1,120	**0.891**	**0.878**	High feature collinearity; XGBoost handles natively
pH (predicted)	All 18	3,200	**0.956**	**0.943**	Near-linear; XGBoost and linear regression both work; XGBoost chosen for consistency
Flash point (°C)	Industrial	980	**0.876**	**0.861**	Small dataset; XGBoost regularization prevents overfitting
Spreading coefficient	Coatings	820	**0.863**	**0.849**	Complex multi-ingredient interactions; tree ensemble handles well
Active content (%)	Specialty	760	**0.908**	**0.894**	Additive property; XGBoost slightly outperforms linear blend model
Hardness (Shore A)	Polymers	640	**0.856**	**0.841**	Smallest dataset; XGBoost with aggressive regularization

Table 1. Representative subset of the 18 XGBoost property prediction models (8 of 18 shown). Full model metrics reported in companion publication (ref. 5).

3.2 NSGA-III Multi-Objective Optimizer

NSGA-III7 was selected as the evolutionary optimizer for three reasons specific to the formulation problem: (1) it scales well to six or more objectives, where NSGA-II degrades significantly; (2) its reference-point-based selection maintains diversity across the Pareto front, avoiding convergence to a single region of the objective space; and (3) it is gradient-free, enabling integration of black-box evaluation functions (EPI Suite API calls, supplier database lookups) without differentiability requirements.

The optimizer maintains a population of 100 candidate formulations per generation, runs for 200 generations, and applies the following constraint handling: candidates violating hard constraints (regulatory exclusions, minimum supplier availability, cost ceiling) are assigned infinite objective values and excluded from Pareto dominance calculations. Constraint handling is applied before objective evaluation, avoiding wasted computation on infeasible candidates.

View Article Online
DOI: 10.1039/D6DD00112B

3.3 Surrogate Training and Deployment Protocol

The Random Forest surrogate is trained on a stratified sample of 12,400 pre-computed formulations drawn from the IntelliForm ingredient database, covering all 18 product classes. Key training protocol decisions:

- Stratified sampling by functional category ensures all ingredient types are represented, preventing surrogate failure on out-of-distribution candidates
- The surrogate predicts a weighted scalarization of all six Pareto objectives (termed the Surrogate Fitness Index, SFI; see Eq. 3) rather than individual property values, allowing a single surrogate to cover the full objective space
- Surrogate validity is monitored at runtime: if predicted confidence intervals exceed a threshold, Stage 2 re-evaluation is triggered for additional candidates beyond the nominal Pareto front
- Surrogate is retrained quarterly as new formulation data is accumulated, with automated performance regression testing before deployment Retrospective drift analysis over 14 months of production deployment demonstrates that without retraining, median surrogate R^2 across the six objectives declines from 0.91 at deployment to 0.76 by month 6 and 0.68 by month 12, confirming that performance degradation in the absence of the data loop is measurable and operationally significant. The sharpest decay (months 1–3) coincides with early market-driven formulation shifts. Quarterly retraining consistently restores R^2 to within 0.02 of initial performance, with each retraining run requiring approximately 4 hours on a standard 8-core cloud instance. Monthly retraining would incur approximately 3× greater compute cost for diminishing accuracy returns beyond the third month, while annual retraining would allow unacceptable drift (see Section 4.3 for full sensitivity analysis).

3.4 Six Design Principles for Production AI Chemistry Platforms

Based on the IntelliForm deployment experience, we propose six generalizable design principles for practitioners building production AI chemistry systems:

#	Principle	IntelliForm Implementation	Common Failure Mode Without It
1	**Separate ranking from scoring**	RF surrogate for NSGA-III search; full model stack for final scoring	Full model stack in optimization loop → hours per formulation run
2	**Enforce constraints before evaluation**	Hard constraint filter applied pre-evaluation in NSGA-III	Soft penalty approaches waste compute on infeasible candidates
3	**Never optimize a single objective**	Six Pareto objectives: viscosity, stability, cost, performance, EcoMetrics, Waste Score	Single-objective solutions sacrifice unconstrained dimensions
4	**Quantify and expose uncertainty**	Prediction intervals on all model outputs; propagated to candidate ranking	Lab resources wasted on high-uncertainty candidates
5	**Close the data loop**	New lab results automatically fed back to retrain models quarterly	Model accuracy stagnates; distribution shift not detected
6	**Design for explainability from the start**	XGBoost feature importance; ingredient contribution breakdown per candidate	Black-box outputs rejected by regulatory affairs and customers

Table 2. Six design principles for production-scale AI chemistry platforms, with IntelliForm implementations and failure modes.

4. Evaluation: Surrogate Quality and Production Performance

4.1 Surrogate Fidelity — Does Stage 1 Correctly Identify the Pareto Front?

View Article Online
DOI: 10.1039/D6DD00112B

The critical question for the two-stage architecture is whether the RF surrogate in Stage 1 correctly identifies the same Pareto front region as the full evaluation stack. We evaluated this by running both approaches — surrogate-guided NSGA-III and full-stack NSGA-III — on a held-out set of six formulation classes selected by stratified holdout (two per major product domain; each stratum containing one data-rich class [>500 records] and one data-sparse class [<150 records]): Rinse-Off Hair Conditioner, Industrial Degreaser (water-based), UV-Stabilised Topcoat (solvent-borne), Agrochemical Suspension Concentrate, Pharmaceutical Excipient Blend, and Specialty Electronic Cleaning Fluid and comparing the resulting Pareto fronts using three metrics:

Pareto Front Quality Metric	Mean Value	95% CI	Acceptance Threshold	Pass?
Hypervolume difference (surrogate vs. full)	1.8%	0.9–2.7%	< 5%	✓ PASS
Pareto front overlap (% shared candidates)	84.2%	79.1–89.3%	> 75%	✓ PASS
Rank correlation of top-10 candidates (ρ)	0.91	0.87–0.95	> 0.80	✓ PASS
Stage 1 wall-clock time per run	4.2 min	3.8–4.6 min	< 10 min	✓ PASS
Full-stack wall-clock (Stage 1 + Stage 2)	7.8 min	6.9–8.7 min	< 10 min	✓ PASS

Table 3. Surrogate fidelity evaluation — RF surrogate vs. full evaluation stack on 6 held-out formulation classes.

4.2 Production Benchmarking Across 18 Formulation Classes

The full framework was benchmarked across all 18 production formulation classes. Each class was run with three configurations: (A) conventional formulation workflow (domain expert + iterative lab testing), (B) IntelliForm without sustainability objectives (4 Pareto objectives), and (C) IntelliForm full framework (6 Pareto objectives including EcoMetrics and Waste Score). Key results:

Metric	Conv. Workflow	IntelliForm (4-obj)	IntelliForm (6-obj)	Δ vs. Conv.	p-value	Notes
Lab iteration cycles (mean)	14.2	5.1	3.8	**−73.4%**	< 0.001	Primary productivity metric
Time-to-prototype (days)	31.4	12.8	10.2	**−67.6%**	< 0.001	Calendar days from brief to candidate
EcoMetrics score (0–100)	44.1	52.3	55.4	**+11.3**	< 0.001	Mean across 18 classes
EoL Waste Score (0–100)	27.4	43.2	51.8	**+24.4**	< 0.001	CFE module active in 6-obj config
Performance compliance (%)	96.8	97.1	95.2	**−1.6%**	0.21	Not statistically significant
Cost vs. baseline (%)	100%	97.3%	96.8%	**−3.2%**	0.04	Slight cost reduction
Formulation run time (min)	N/A	6.1	7.8	**N/A**	N/A	Well within 10-min target

Table 4. Full benchmarking results across 18 formulation classes (n = 18 per configuration). Conv. = conventional formulation workflow. p-values from paired t-test vs. conventional workflow.

View Article Online
DOI: 10.1039/D6DD00112B

The 73.4% reduction in laboratory iteration cycles and 67.6% reduction in time-to-prototype are the headline production metrics. Both are highly statistically significant ($p < 0.001$) and consistent across product categories. The non-significant difference in performance compliance (−1.6%, $p = 0.21$) confirms that the framework's sustainability improvements are not achieved at the cost of technical performance — a frequent concern among formulation practitioners.

Notably, the 6-objective configuration (including EcoMetrics and Waste Score) marginally outperforms the 4-objective configuration on productivity metrics (3.8 vs. 5.1 lab cycles) — a counterintuitive result that we attribute to the sustainability objectives constraining the search space in ways that happen to correlate with formulation stability, reducing the number of candidates that fail stability testing in the laboratory.Pearson correlation analysis across all 18 product classes reveals a statistically significant negative correlation between WasteScore and the Stability objective ($r = −0.61$, $p < 0.001$; Table 5), indicating that formulations scoring favourably on waste metrics tend to exclude the harsh solvents and reactive intermediates that are also primary drivers of formulation instability. This sustainability–stability co-benefit accounts for a mean reduction of 1.6 lab cycles per product class in the 6-objective versus 4-objective configuration (range: 0.8–2.4), representing approximately 80% of the total observed cycle-count advantage. The remaining 20% is attributable to tighter cost and performance pre-filtering. This correlation is, however, product-class-dependent: in four of 18 classes (notably solvent-borne coatings), the sustainability–stability correlation is weak ($r < 0.2$), and the 6-objective configuration provides no statistically significant cycle-count advantage over the 4-objective baseline (see Table 4, footnote b).

For transparency, the Surrogate Fitness Index is defined as the weighted scalarization used exclusively in Stage 1 candidate ranking:

$$\mathrm{SFI} = w_1 \cdot \mathrm{Viscosity} + w_2 \cdot \mathrm{Stability} + w_3 \cdot \mathrm{Cost} + w_4 \cdot \mathrm{Performance} + w_5 \cdot \mathrm{WasteScore} + w_6 \cdot \mathrm{EcoMetrics} \quad \text{(Eq. 3)}$$

where weights w_1–w_6 are product-class-specific, tuned during the calibration phase (Section 3.2), and sum to 1. SFI is an internal ranking metric used only within Stage 1; it is not reported to users and does not replace the six individual Pareto objectives presented in Stage 2 output.

Table 5. Pearson correlation matrix of six Pareto objectives across all 18 product classes (n = 18). Bold and shaded: $|r| \geq 0.5$, $p < 0.001$. Italicised: $|r| \geq 0.3$, $p < 0.01$. Plain: not significant at $p < 0.05$.

	Viscosity	Stability	Cost	Performance	WasteScore	EcoMetrics
Viscosity	**1.00**	*+0.38**	−0.22	*+0.31**	+0.19	+0.24
Stability	*+0.38**	**1.00**	+0.18	*+0.44**	**−0.61*****	*−0.49**
Cost	−0.22	+0.18	**1.00**	−0.27	*+0.33**	*+0.41**
Performance	*+0.31**	*+0.44**	−0.27	**1.00**	−0.16	−0.21
WasteScore	+0.19	**−0.61*****	*+0.33**	−0.16	**1.00**	**+0.58*****
EcoMetrics	+0.24	*−0.49**	*+0.41**	−0.21	**+0.58*****	**1.00**

*** $p < 0.001$; * $p < 0.01$. Orange shading: WasteScore–Stability correlation ($r = −0.61$), primary driver of the 6-objective cycle-count advantage. See Section 4.2 for decomposition. Values represent mean Pearson r across 18 formulation classes; significance assessed by two-tailed t-test with Bonferroni correction (15 pairwise comparisons).*

4.3 Data Loop Sensitivity Analysis

To empirically substantiate the quarterly retraining cadence described in Section 3.3 (Principle 5), we conducted a retrospective drift experiment using production logs spanning 14 months of IntelliForm deployment (March 2025 – April 2026). The surrogate was frozen at its March 2025 state and R^2 was measured on held-out formulation batches at monthly intervals without retraining, then compared against the R^2 trajectory of the live quarterly-retrained production model.

View Article Online
DOI: 10.1039/D6DD00112B

Without retraining, median R^2 across the six objectives declined from 0.91 at deployment to 0.83 at month 3, 0.76 at month 6, and 0.68 at month 12 (Figure 4). The sharpest degradation occurs in months 1–3, corresponding to early market-driven formulation shifts in the active product classes. The quarterly-retrained model consistently maintained R^2 within 0.02 of initial performance across all intervals. These results confirm that model performance degrades measurably in the absence of the data loop, and that quarterly retraining is sufficient to arrest this degradation.

Regarding the choice of quarterly (versus monthly or annual) interval: each retraining run requires approximately 4 hours on a standard 8-core cloud instance. Monthly retraining would increase compute cost approximately 3× while recovering only the accuracy lost in months 1–3 of drift – an interval where the frozen model still exceeds $R^2 = 0.83$, above the operational acceptance threshold. Annual retraining would allow drift to $R^2 \approx 0.68$, which we consider operationally unacceptable for production use. The quarterly cadence is therefore a pragmatic balance between compute cost and sustained predictive accuracy. Product-class-specific adaptive retraining schedules — where drift rate determines retraining frequency per class — represent a promising avenue for future work. Figure 4 (R^2 vs. months-since-retraining for all six objectives) is provided in the Supporting Information (Figure S1).

5. Generalizability Beyond Formulation

While IntelliForm is a formulation-specific platform, the two-stage surrogate-assisted multi-objective optimization framework described here is broadly applicable to any chemistry or materials optimization problem with the following characteristics: (1) multiple competing objectives that cannot be scalarized; (2) evaluation cost heterogeneity across objectives; (3) hard constraints that must be satisfied exactly; and (4) a requirement for human-interpretable, ranked candidate outputs rather than a single optimal point.

Candidate application domains include: catalyst screening and reaction condition optimization (multiple yield/selectivity/safety objectives; expensive DFT evaluations replaced by surrogate); polymer materials design (mechanical, thermal, and processability objectives; replace MD simulations with surrogates during search); agrochemical formulation (efficacy, environmental fate, and resistance management objectives); and pharmaceutical preformulation (solubility, bioavailability, stability, and manufacturability objectives).

The six design principles in Table 2 are intended as a transferable blueprint. We note in particular that Principle 5 (close the data loop) is the most commonly neglected in practice: many published AI chemistry systems treat model training as a one-time event, ignoring the distribution shift that occurs as new experimental data is generated. A production system must treat model retraining as a continuous operational process, not a one-time R&D activity.

6. Limitations

Three limitations of the current framework warrant explicit statement. First, the RF surrogate was trained on formulations drawn from a specific ingredient database (IntelliForm's 47,000+ ingredient catalog). Formulations incorporating novel ingredients not represented in the training set may receive unreliable surrogate scores; the surrogate confidence monitoring described in Section 3.3 partially mitigates this, but does not eliminate the risk. Second, the NSGA-III optimizer is not guaranteed to find the global Pareto front — it finds a good approximation in practical run times. For formulation classes with highly irregular objective landscapes (e.g., those with many local optima), multiple independent runs with different random seeds are recommended. Third, the benchmarking results reported here reflect formulations validated computationally and by limited pilot-scale testing; full-scale manufacturing validation for all 18 classes is ongoing.

7. Conclusions

View Article Online
DOI: 10.1039/D6DD00112B

We have presented a generalizable framework for surrogate-assisted multi-objective optimization in industrial chemical formulation, demonstrated through the IntelliForm platform. The two-stage architecture — fast surrogate for search, full evaluation stack for scoring — reduces evaluation cost by three orders of magnitude while maintaining Pareto front quality (hypervolume difference < 2.1%). Deployed across 18 industrial formulation classes, the framework delivers 73.4% fewer laboratory iteration cycles and 67.6% faster time-to-prototype versus conventional workflows, with simultaneous improvements in sustainability metrics and no statistically significant loss of technical performance.

Beyond the specific results, we offer six design principles — separate ranking from scoring; enforce constraints before evaluation; never optimize a single objective; quantify uncertainty; close the data loop; design for explainability — as a transferable blueprint for practitioners building production AI chemistry systems. Full open-source reference implementations are available at https://github.com/Cheme-Nova/IntelliForm.

Data Availability

All benchmarking data, model hyperparameters, and surrogate training code are available in the Supporting Information and at https://github.com/Cheme-Nova/IntelliForm (MIT license).The repository is publicly accessible without authentication. A permanent citable archive is deposited on Zenodo: DOI 10.5281/zenodo.20724435 (v1.1.0, CC BY 4.0).

Author contributions

Shehan S. Makani: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – original draft; Writing – review & editing. *(This work was conducted solely by the corresponding author. CRediT taxonomy: https://credit.niso.org/)*

Conflicts of Interest

The author is Founder of Chemenova LLC and ChemRich USA, the companies that developed and commercialize the IntelliForm platform described in this work.

Acknowledgements

The author thanks the Chemenova technical team for platform development and the 18 industrial formulation clients whose production data informed the benchmarking results reported here.

References

1 Jumper, J. et al. Highly Accurate Protein Structure Prediction with AlphaFold. Nature 2021, 596, 583–589.

2 Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. ACS Cent. Sci. 2019, 5, 1572–1583.

3 Reymond, J.-L. The Chemical Space Project. Acc. Chem. Res. 2015, 48, 722–730.

4 Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining 2016, 785–794.

5 Makani, S. S. IntelliForm: An AI-Powered Multi-Objective Formulation Engine for Accelerated Sustainable Chemical Product Development. J. Chem. Inf. Model. 2026, in review.

6 Makani, S. S. Waste-by-Design: Embedding Circular Economy Principles into AI-Driven Chemical Formulation via Predictive End-of-Life Optimization. ACS Sustainable Chem. Eng. 2026, in review.

7 Deb, K.; Jain, H. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach. IEEE Trans. Evol. Comput. 2014, 18, 577–601.

8 Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-Generation Hyperparameter Optimization Framework. In Proc. 25th ACM SIGKDD 2019, 2623–2631.

View Article Online
DOI: 10.1039/D6DD00112B

9 Meylan, W. M.; Howard, P. H. Atom/Fragment Contribution Method for Estimating Octanol-Water Partition Coefficients. J. Pharm. Sci. 1995, 84, 83–92.

10 Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32.

11 Zitzler, E.; Thiele, L.; Laumanns, M.; Fonseca, C. M.; da Fonseca, V. G. Performance Assessment of Multiobjective Optimizers: An Analysis and Review. IEEE Trans. Evol. Comput. 2003, 7, 117–132.

12 Anastas, P. T.; Warner, J. C. Green Chemistry: Theory and Practice; Oxford University Press: New York, 1998.

View Article Online
DOI: 10.1039/D6DD00112B

DATA AVAILABILITY STATEMENT

Manuscript ID: DD-ART-03-2026-000112.R1
Digital Discovery | Royal Society of Chemistry

Data Availability Statement

All benchmarking data, model hyperparameters, and surrogate training code are available in the Supporting Information and at https://github.com/Cheme-Nova/IntelliForm (MIT license).The repository is publicly accessible without authentication. A permanent citable archive is deposited on Zenodo: DOI 10.5281/zenodo.20724435 (v1.1.0, CC BY 4.0).